基于多通路融合网络的高速公路雾天能见度等级识别 *

闫宏艳[1]，孙玉宝[1†]，张振东[2]，黄　亮[2]

(1. 南京信息工程大学 江苏省大数据分析技术重点实验室 江苏省大气环境与装备技术协同创新中心，南京 210044; 2. 江苏省气象局 江苏省气象服务中心 中国气象局交通气象重点开放实验室，南京 210008)

摘　要：雾天是影响高速公路交通安全的重要因素。研究从监控图像进行高速公路雾天能见度的自动识别方法可以为交通管理部门的智能管理和决策提供技术支持。根据大气散射模型分析出与雾浓度相关的多个物理因素，提出了综合该些物理因素的多通路融合识别网络。该网络使用三个通路联合学习深度视觉特征、传输矩阵特征和场景深度特征，并设计注意力融合模块来自适应地融合这三类特征以进行能见度等级识别。同时构建了一个合成数据集和一个真实的高速公路场景数据集，用于网络参数学习和性能评估。实景数据集中的图像是从中国多条高速公路的监控视频中收集的。在这两个数据集上的实验表明，所提出的方法可以适应不同的监控拍摄场景，能够比现有方法更准确地识别能见度等级，有效提升了识别精度。

关键词：能见度识别；多通路网络；大气散射模型；注意力融合

中图分类号：TP751　　doi: 10.19734/j.issn.1001-3695.2022.01.0010

Recognition of highway visibility level in foggy weather using multi-stream deep fusion network

Yan Hongyan[1], Sun Yubao[1†], Zhang Zhendong[2], Huang Liang[2]

(1. *Jiangsu Key Laboratory of Big Data Analysis Technology* (*B-DAT Lab*) *Jiangsu Collaborative Innovation Center on Atmospheric Environment & Equipment Technology*, *Nanjing* 210044, *China*; 2. *Jiangsu Meteorological Service Center*, *Jiangsu Provincial Meteorological Bureau*, *China Meteorological Administration Transportation Meteorology Key Laboratory*, *Nanjing* 210008, *China*)

Abstract: Foggy weather is an important factor affecting highway traffic safety. Research on the automatic recognition method of highway fog visibility from surveillance images can provide technical support for the intelligent management and decision-making of the traffic management department. This paper analyzes multiple physical factors related to fog density based on the atmospheric scattering model and proposes a multi-channel fusion network that integrates these physical factors. Specifically, the method jointly exploits three streams to learn deep visual feature, transmission matrix feature and scene depth feature, and designs an attention fusion module to adaptively fuse these three streams for the final visibility level recognition, which is very beneficial for improving the recognition accuracy. Meanwhile, this paper constructs a synthetic dataset and a real-scene dataset for network parameters learning and performance evaluation. The images in the real-scene dataset are collected from surveillance videos of multiple highways in China. Experiments on these two datasets show that this method can identify visibility level more accurately than existing methods.

Key words: visibility level recognition; multi-stream network; atmospheric scattering model; attention fusion

0　引言

雾是自然环境中常见的天气现象，雾天条件下光线被空气中的悬浮颗粒吸收和散射，导致能见度距离降低。这会对现实生活的诸多方面造成潜在的安全隐患，尤其是在公路运输中，能见度的降低会大大增加交通事故的发生率。据统计，影响道路正常运行的恶劣天气中大雾为主要影响天气，占比达 69%。雾天时，发生交通事故的概率会比晴天时高出几十倍。党中央高度重视交通安全问题，各级部门多次提出相关指导意见和规划纲要。《国务院关于加强道路交通安全工作的意见》中也指出，要加强道路交通安全设施建设，积极推进高速公路灾害性天气预报和预警系统建设，提高对浓雾等恶劣天气的防范应对能力。因此迫切需要建立高速公路雾天能见度自动识别系统，这样高速公路管理部门能够获得准确和即时的决策支持，进而采取相应的控制措施，对降低事故率具有很大帮助[1]。此外，由于能见度下降会导致自动驾驶辅助系统的失效，因此这些基于视觉的自动驾驶系统也需要能见度识别，从而可以及时调整操作或对驾驶员作出提醒[2]。

目前广泛应用的能见度识别方法主要包括人眼观测和仪器测量。人眼观测是一种主观的估计方法，不仅耗费人力而且存在较大误差。仪器测量主要是通过放置在室外的能见度传感器来测量能见度距离。虽然仪器测量提高了识别精度，但仪器部署成本高，识别范围非常有限，难以实现大规模覆盖。目前的高速公路通常都配备了良好的监控系统。图 1 显示了大雾天气下不同能见度范围的高速公路监控图像。根据不同雾浓度下监控图像的差异性，可以利用计算机视觉和深度学习方法，建立一种成本低、灵活性高的自动识别模型。

为了提升识别算法的适应性，本文着重研究仅使用单张高速公路雾天图像的能见度等级识别方法。但由于监控摄像的视角和光照条件的不同，准确识别不同场景下雾天图像的能见度等级是一项具有挑战性的任务。现有的研究大多集中

收稿日期：2022-01-07；**修回日期：**2022-02-28　**基金项目：**国家自然科学基金资助项目(U2001211)；江苏省气象局重点基金资助项目(KZ202105)

作者简介：闫宏艳(1997-)，女，河北保定人，硕士研究生，主要研究方向为深度学习、能见度检测；孙玉宝(1983-)，男(通信作者)，江苏连云港人，教授，博士，主要研究方向为深度学习理论与方法、计算机视觉(sunyb@nuist.edu.cn)；张振东(1988-)，男，江苏南通人，高级工程师，硕士，主要研究方向为交通场景图像分析；黄亮(1981-)，男，湖南常德人，高级工程师，副主任，学士，主要研究方向为交通气象．

在图像去雾算法[3~6]上，然而它们是两个不同的任务。图像去雾是一种通过调整每个像素的值来提高图像质量的图像增强任务，能见度识别则是通过分析图像的雾分布来推断能见度水平。目前，与图像去雾算法相比，从雾天图像中识别能见度的工作较少，大致可以分为两类：基于物理模型的方法和基于深度学习的方法。基于物理模型的方法主要是基于大气散射模型，它首先估计大气散射模型中的场景深度参数和传输矩阵参数，然后根据一些启发式规则来估计能见度。虽然这种方法有较好的物理解释性，但并不能很好地适应高速公路的实际复杂场景。基于深度学习的方法主要利用卷积神经网络优秀的学习能力，直接学习从雾天图像到能见度水平的函数映射。但由于监控摄像机视角和光照条件等因素，视觉特征会有显著差异，进而影响识别的准确性。





图 1　不同雾浓度下高速公路监控图像

Fig. 1　Some highway surveillance images under different foggy densities

为了应对上述问题，本文根据雾形成的物理模型，提出了高速公路雾天图像能见度识别的多通路深度融合网络模型，称为 MSVP-Net。该方法融合了传输矩阵特征、场景深度特征和视觉特征进行等级识别。具体地，首先传输矩阵通路通过暗通道先验算法估计雾天图像的传输矩阵、场景深度通路通过深度估计子网络估计雾天图像深度图、视觉特征通路通过残差子网络提取图像的深度视觉特征，然后设计了注意力融合模块自适应地融合这三个通路的特征，最后通过全连接层对融合的特征进行最终的能见度等级分类。为了训练和评估所提出的网络，本文构建了两个数据集，一个是合成数据集，另一个是真实的高速公路场景数据集。合成数据集使用 FRIDA1 和 FRIDA2 中的无雾图像及其深度图来生成不同强度的有雾图像，真实场景数据集是从我国的多个高速公路监控系统中收集到的图像，这些监控图像的能见度等级是由专业气象人员进行标定。在这两个数据集上的实验表明，本文提出的方法比现有方法更能准确地识别高速公路能见度。本文的主要贡献如下：

1)提出了一种新的网络模型，仅从单张雾天监控图像识别高速公路能见度等级。该方法可为高速公路交管部门提供低成本且高效的智能管控技术支持。

2)本文根据大气散射模型提出了由深度视觉特征通路、传输矩阵通路和场景深度通路组成的多通路网络，并通过注意力融合模块对三个通路进行自适应融合，进而识别能见度等级，有效提升了识别精度。

3)本文构建了一个合成数据集和一个真实场景的高速公路数据集进行性能评估，在这两个数据集上的实验结果验证了该网络的优越性能。

1　相关工作

在本节中，首先阐述了雾天成像的大气散射模型，然后对现有基于图像的能见度识别工作进行了综述。这些现有工作大致可以分为两类，一类是基于物理模型的方法，另一种是基于深度学习的方法。

1.1　大气散射模型

1999 年，Srinivasa G. Narasimhan 等人提出了大气散射模型，定量建模雾天成像过程[7]。此模型假设成像设备所接收到的光强来自两个部分：一部分是由物体反射，经由大气粒子衰减最终到达成像设备的光强；另一部分主要是由大气介质中的太阳光散射形成的大气光强。其模型的具体形式可以表示：

$$I(x)=J(x)t(x)+A(x)(1-t(x)) \tag{1}$$

其中，x 为图像中像素的空间坐标，$I(x)$ 为拍摄到的雾天图像，$J(x)$ 为对应的无雾图像。$A(x)$ 为大气环境中的光强，通常将它假设为一个常数变量。$t(x)$ 为传输矩阵，其物理含义是由目标物体反射的光经大气粒子散射后能够达到成像设备的能力。对于 RGB 图像，一般认为三色通道具有相同的传输矩阵。当假设大气光均匀时，$t(x)$ 可以表示如下：

$$t(x)=e^{-\beta d(x)} \tag{2}$$

其中，β 为大气衰减系数，$d(x)$ 为像素 x 的场景深度。

1.2　基于物理模型的图像能见度识别方法

许多传统的图像能见度识别方法[8~14]主要是基于传统图像处理或大气散射模型。文献[11]依据 Koschmieder 定律指出，光会受漂浮在空中的水粒子影响而发生扩散。基于这一现象，Hautiere 等人发现地平线上的灰度变化可以用来测量图像中的雾浓度，进而估计能见度距离。在文献[12]中，R.Gallen 等人基于判别外部光源周围是否存在后向散射的光晕来检测是否有雾，并由此估计气象能见度距离。在文献[13]中，S.Bronte 等人进一步指出，由于水颗粒和其他散射介质漂浮在空气中，来自天空的光会扩散并聚焦在道路区域，道路与天空之间的边界会变得模糊，因此可以通过计算相对天空高度得到能见度距离。在文献[14]中，J.Mao 等人依据大气散射模型构造了一个可调的经验函数来识别雾霾程度，并对多种户外图像进行了统计分析，通过多元线性回归估计了经验函数中的超参数。然而，这种启发式设计的经验函数并不能很好地推广到实际的高速公路场景中，影响了识别性能。

1.3　基于深度学习的图像能见度识别方法

近年来，卷积神经网络在图像识别[15,16]、目标检测[17,18]等计算机视觉任务中都取得了巨大的成功，并被应用于图像去雾任务。然而，图像能见度识别与图像去雾是两个不同的任务，目前只有少数工作尝试使用卷积神经网络来识别能见度水平。Zhang J 等人提出了一种可用于检测室外图像雾霾浓度的端到端卷积神经网络。具体而言，该方法使用了雾霾图像和相同场景下的无雾图像共同引导网络训练，将两张图像中的结构相似性(SSIM)分数作为回归目标[5]。然而，在实际应用中较难收集无雾和相应有雾的图像对。Gunawan 等人使用经典的 Alexnet 从有雾图像中检测能见度信息[19]。Li Y 等人提出了一种估计和监测城市空气污染的方法，该方法利用暗通道先验[3]估计传输矩阵，利用深度卷积神经网络[20]估计深度图，并设计预定义的转换和池化函数的组合来识别雾霾水平[21]。但是，预定义的变换函数限制了深度网络的表示能力，不利于提高识别精度。总结来说，从单张雾天图像识别能见度水平是一个具有挑战性的问题，如何更好地融合多种类型的特征是应对这一挑战的关键。

2　多通路融合识别网络

由大气散射模型可知，传输矩阵与场景深度信息是估计雾浓度的两个重要参数。同时，学习视觉特征也有利于分析雾的分布。为此，本文提出了大气散射模型启发的多通路深度融合识别网络。该网络采用三条通路联合学习输入图像的传输矩阵、场景深度和视觉特征，并通过注意力模块自适应融合三条通路特征进行高速公路雾天图像能见度等级识别。在本节中，将首先介绍所提出方法的总体结构，然后阐述注意力融合模块设计，最后介绍网络参数的学习方法。

2.1　网络架构

该网络的总体架构如图 2 所示。首先对输入的雾天场景

图像进行三通路并行处理，分别得到估计的场景深度、传输矩阵和深度视觉特征。然后将三部分特征输入到注意力融合模块中进行自适应融合，由全连接层对融合特征进行分类，从而识别出能见度等级。

2.1.1 场景深度通路

该通路用于提取输入图像的场景深度特征。场景深度特征反映了场景中的物体与监控摄像机之间的距离，这是估计能见度水平的一个重要依据。本文选用 FastDepth 网络[22]对输入图像进行景深估计。FastDepth 网络采用编码器-解码器结构进行单目深度估计，其显著优点是结构简单、模型轻量化，同时可以保持较高的精度。为了降低计算复杂度，网络编码器采用 MobileNet 模型，将编码器中的标准卷积层分解为深度卷积和逐点卷积。解码器执行 5 次上采样，中间三次上采样的结果通过 Skip Connections 的方法分别与编码器部分的特征进行了特征融合，为了减小上采样部分的通道特征，还使用了 5×5 的卷积来降维。最后使用 1×1 的卷积得到高分辨率深度图。

2.1.2 传输矩阵通路

此通路用于估计输入雾天图像的传输矩阵。根据式(2)，传输矩阵是大气消光系数与场景深度乘积的负指数函数。通过估计传输矩阵，可以来推断雾气浓度。根据暗通道先验模型[3]，使用下面公式估计传输矩阵 $\tilde{t}$：

$$\tilde{t} = 1-\omega \min_{y\in\Omega(x)}\left(\min_{c} \frac{J^c(x)}{A^c}\right) \tag{3}$$

其中，c 为颜色通道，$\omega(0<\omega<1)$ 是一个常数参数，$\Omega(x)$ 是一个以 x 为中心的图像块，$J^c(x)$ 是通道 c 在 x 位置的像素值，A^c 是估计的天空亮度。在本文实验中，ω 设为 0.95，图像块的数量设置为 5。并进一步使用导向滤波对估计的传输矩阵 $\tilde{t}$ 进行细化。

2.1.3 深度视觉特征通路

该通路是由多个残差块组成的卷积网络，用于从输入的图像中学习深度视觉特征。残差模块由于其优越的学习能力[16]被广泛应用于计算机视觉任务中。与文献[16]类似，所使用的残差块为三层结构。第一层和第三层都使用 1×1 卷积，分别用于压缩特征、扩大特征映射的通道维数。通过这种瓶颈设计，第二层卷积层只需要处理低通道维度的特征。对于雾天能见度识别任务，还需要提取多尺度特征来更好地学习雾分布。因此，本网络将残差块的第二层设计为三个并行的空洞卷积，其空洞率分别为 1、3 和 5，从而能够提供不同感受野范围内的多尺度特征。为了减低特征图的空间分辨率，首先在三个残差结构之前进行了步幅为 2 的 3×3 卷积和最大池化操作。同时，将其中两个残差块中的卷积步幅设置为 2。该通路共执行了四次 ×2 降采样。

2.2 注意力融合模块

本文设计了一个基于注意力机制的自适应融合模块，对三个通路输出的场景深度 $d(x)$、传输矩阵 $t(x)$ 和深度视觉特征 F 进行有效融合。多通路注意力融合模块的整体结构如图 3 所示。考虑到场景深度和传输矩阵是与雾浓度相关的重要物理因素，因此首先通过 concatenate 操作将 $d(x)$ 与 $t(x)$ 进行融合，并通过卷积模块进一步提取高层特征。该卷积模块由两个 3×3 卷积层、两个最大池化层组成，步幅为 2，保证卷积模块的输出和深度视觉特征 F 具有相同的空间分辨率。将卷积模块处理后的特征与深度视觉特征 F 通过 concatenate 操作进行融合，同时使用 1×1 卷积来减少融合后的特征通道数，将其记为 F_s。

特征 F_s 中每个通道和空间位置对能见度水平识别具有不同贡献度。因此，本文通过注意力机制自适应地关注空间和通道维度上的重要特征。注意力融合模块配置为残差空间注意和残差通道注意的级联结构，残差注意力的优点是能够捕获信息特征，同时减少信息损失和学习困难。

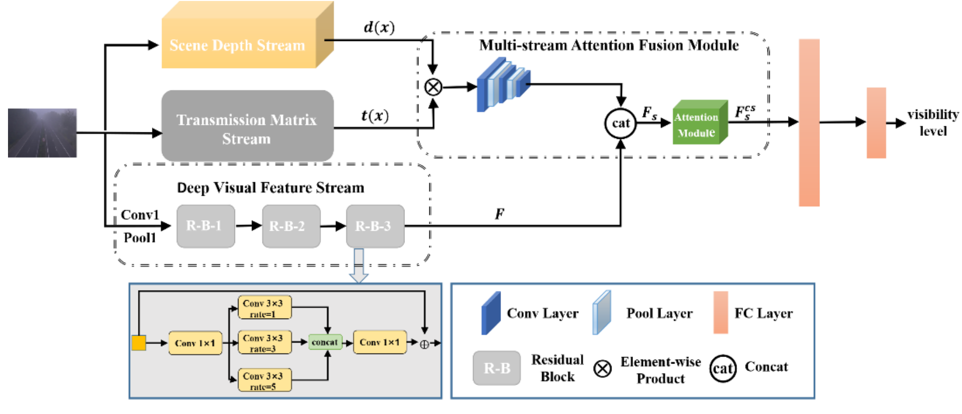


图 2 大气散射模型启发的雾天高速公路能见度识别网络

Fig. 2 Highway visibility recognition network in foggy weather inspired by atmospheric scattering model

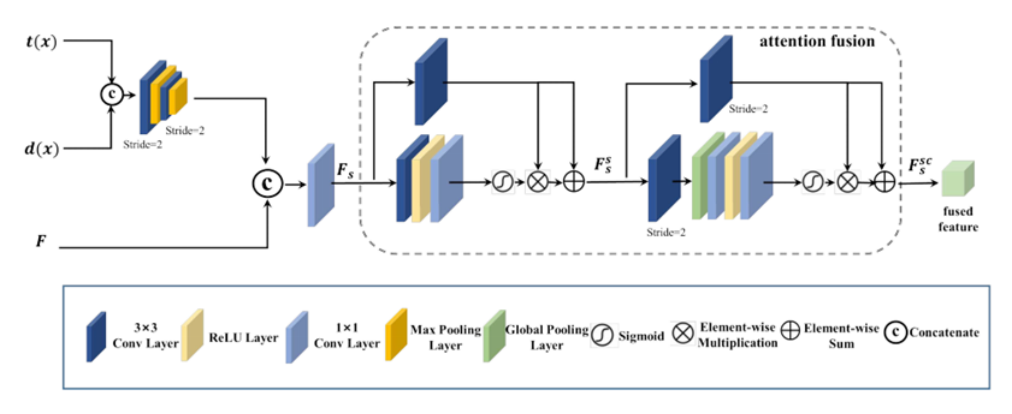


图 3 注意力融合模块结构图

Fig. 3 The detailed architecture of attention fusion module

空间注意力通过分配不同的权重来突出对能见度识别任务重要的空间位置。残差空间注意力的详细计算可以表示为

$$S_a = \sigma(Conv1(\delta(Conv3(F_s)))) \tag{4}$$

$$F_S^s = (1+S_a)\otimes Conv3(F_s) \tag{5}$$

其中 $Conv3$ 为 3×3 卷积操作，$Conv1$ 为 1×1 卷积操作，δ 为 ReLU 激活函数，σ 为 sigmoid 激活函数，$\otimes$ 为元素乘法。首先通过卷积操作得到维度大小为 $1\times H\times W$ 的空间权重 s_a，然后依据式(5)计算得到空间注意加权特征 F_S^s。

类似地，通道注意力主要通过分配不同的权重来突出重要的特征通道。首先，对空间注意力模块得到的 F_S^s 进行全局平均池化操作 H_p，具体为

$$f_c = H_p(F_s) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{Sc}^s(i,j) \tag{6}$$

其中 $F_{Sc}^s(i,j)$ 为 F_S^s 在位置 (i,j) 上的第 C 通道的值。H_p 操作将特征维度从 $C\times H\times W$ 转换为 $C\times 1\times 1$，然后利用所得到的特征 f_c 来估计通道注意权重 C_a，并相应地计算出通道注意加权特征 F_S^{SC}。具体地计算公式可以表示为

$$C_a = \sigma(Conv1(\sigma(Conv1(Conv3(f_c))))) \tag{7}$$

$$F_S^{SC} = (1+C_a)\otimes Conv3(F_S^s) \tag{8}$$

本文将上述 $Conv3$ 操作的步幅设置为 2，以减小特征图的空间维度大小。F_S^{SC} 是残差注意力处理后的最终融合特征，将其输入到全连接层中进行最终能见度等级分类。

2.3　损失函数

该网络包含多个需要优化的参数块，即场景深度通路、视觉特征通路、多通路注意力融合模块和全连接层。为了减少学习难度，本文采用两阶段化的学习方法来更新网络参数。在第一阶段，只学习场景深度通路的参数 Θd。在第二阶段，固定了场景深度通路的参数 Θd，重点学习剩余的部分，包括深度视觉特征流的参数 Θv、注意融合模块的 Θa 和全连接层的 Θf。对于第二学习阶段，损失函数采用交叉熵函数，具体定义为

$$L(\Theta) = -\sum_{i=1}^{C} y_c \log(F(\Theta, x_i)_c) \tag{9}$$

其中 $\Theta=\{\Theta v,\Theta a,\Theta f\}$ 为网络中需要更新的参数，x_i 为第 i 个训练样本，$F(\Theta,\cdot)$ 为网络整体的相关函数映射，$F(\Theta,x_i)_c$ 表示 x_i 分类为第 c 类的概率。y_c 表示二进制变量，如果 x_i 的类别与其真类别相同，则为 1，否则为 0。C 是能见度级别的数量。完成网络训练后，网络可用来识别新测试样本的能见度等级。

3　实验部分

3.1　数据集介绍

在本工作中，构建了两个数据集进行网络训练与实验评估，包括合成数据和真实场景数据集。

FRIDA 数据集：FRIDA 数据集由 FRIDA1 数据集和 FRIDA2 数据集构建。FIDA1 和 FRIDA2 最初是为图像去雾任务而设计的，FRIDA1 包含 18 个城市道路场景的 90 张合成图像[23]，FRIDA2 包含 66 个不同的道路场景的 330 张合成图像[24]，每个合成场景都配有一张清晰图像和一张相应的场景深度图。根据式(2)，本文使用清晰的图像和相应的场景深度图，通过改变 β 的值来生成多张有雾图像，以模拟不同能见度等级下的图像。图 4 显示了用不同的 β 值生成的一些样本，β 值越大，雾浓度越大。总共生成了 756 张雾图像，并根据能见度将图像分为四个级别：无雾、轻雾、中雾和浓雾。



图 4　通过改变 β 值生成的一些样本图片

Fig. 4　Some images generated by changing the value of　β

高速公路数据集：本文收集了 1200 张真实的高速公路监控图像，这些图像是 2019 年至 2020 年间由中国多条高速公路上的大量监控摄像头拍摄得到的。根据交通管理部门的应用要求，专业的气象学家根据能见度距离将能见度分为四个级别。表 1 给出了详细的能见度级别标准，等级越高，图像就越清晰，能见度距离就越长。该高速公路数据集的真实能见度级别由专业气象学家和交通管理人员共同确定，图 5 显示了来自这四个能见度级别的部分样本图像，每行为同一级别。这些高速公路监控图像的成像条件差异很大，例如相机高度和观看方向、天气条件和照明强度可能都不同。

表 1　能见度等级标准

Tab. 1　Visibility level standard

visibility level	0	1	2	3
visibility distance	0-100m	100-200m	200-500m	500m+

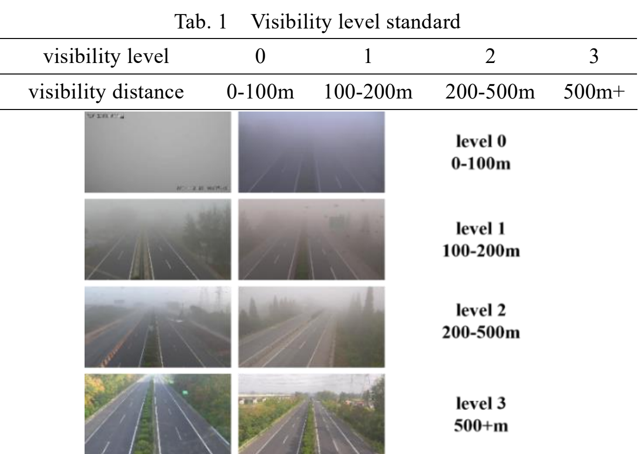


图 5　来自四个级别的部分图像

Fig. 5　Some sample images from four visibility levels

3.2　实验设置

在实验中所提出的网络将分两阶段进行训练。首先使用 NYU Depth V2 数据集[25]对场景深度通路进行预训练。然后固定场景深度通路的参数，更新其余参数,其中传输矩阵通路没有可学习的参数。本文使用 PyTorch 框架搭建网络，使用 GPU2080Ti 训练网络，采用 ADAM 方法[26]作为优化器来更新网络参数。优化器中参数的详细设置包括：$\beta_1=0.9$ 和 $\beta_2=0.9999$。学习速率初始化为 1×10^{-4}，在每 30 个周期后下降到一半，batch-size 的大小设置为 16。

3.3　对比实验

在本文中，将所提出的方法与现有的基于深度学习的能见度识别方法在 FRIDA 数据集和高速公路数据集上进行了对比实验。选取的对比方法包括 AlexNet[27]、depth ⊗ trans[21](深度信息+暗通道先验)和 MSBDN[28]。AlexNet 方法是仅通过 AlexNet 提取可视化特征,用于能见度等级识别[19]。depth ⊗ trans 最初是为估计雾霾水平而设计的回归模型[21]，为了使该模型适应离散的能见度等级分类任务，将其输出层修改为全连接层，而骨干网络保持不变，骨干网络中深度图和传输矩阵的变换函数选取为单位变换 $T(x)=x$，并将这两个特征连接起来进行分类。此外本文还修改了一种基于深度学习的图像去雾算法，通过保留主干网络和添加全连接层来识别能见度级别。在实验中，选择了一个最新的、高效的图像去雾网络 MSBDN[28]，使用其编码器从雾天图像中提取特征，然后利用提取的特征对能见度级别进行识别。对比实验结果如表 2 所示，被加粗的为最好精度。

据表 2 的实验结果，depth ⊗ trans 方法在两个数据集上的性能都较差。主要原因是此方法只结合估计的场景深度和传输矩阵来识别能见度级别，这对传输矩阵和场景深度的估计精度有很高的要求。然而，在真实的场景中，很难准确地估计这些参数。MSBDN 的性能优于 AlexNet，这表明 MSBDN 的编码器网络相比于 AlexNet 可以学习到更多的信息性视觉特征。本文所提出的 MSVP-Net 在两个数据集上都具有最好

的识别精度，特别地，该方法在真实场景高速公路数据集上具有较大的性能优势。图 6 展示了这些方法在高速公路数据集上的识别结果的混淆矩阵，可以看出本文提出的方法更具有分类稳定性。图 7 展示了四组在高速公路数据集上的四种方法的识别结果和真实结果的对比样本，尽管每个场景的监测角度和光照条件存在差异，但本文方法可以对这四个样本场景都作出正确的识别。图 8 展示了本文方法在合成数据集和高速公路数据上的部分识别结果，可以得知该方法在大部分情况下都可以作出准确的识别。图中 1~4 行对应高速公路数据集，第 5 行对应合成数据集。

表 2 四种方法在 FRIDA 数据集和高速公路数据集上的识别精度

Tab. 2 The recognition accuracy of multiple methods upon FRIDA and Highway datasets

dataset	AlexNet[19]	depth ⊗ trans[21]	MSBDN[28]	MSVP-Net
FRIDA	98.59%	93.84%	98.59%	**99.15%**
Highway	71.06%	70.21%	77.45%	**84.68%**

(a) AlexNet (b) depth ⊗ trans

(c) MSBDN (d)MSVP-Net

图 6 四种方法在高速公路上的识别精度混淆矩阵

Fig. 6 The confusion matrix of multiple methods on the highway dataset

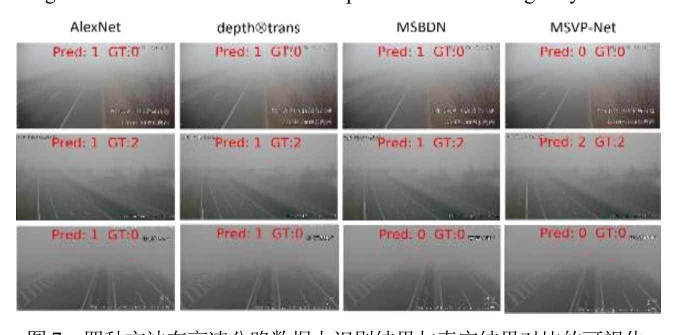


图 7 四种方法在高速公路数据上识别结果与真实结果对比的可视化

Fig. 7 Visualization of the comparison of the four methods between the recognition results and the real results on the highway data



图 8 本文方法在高速公路数据集与合成数据集上识别结果与真实结果对比的可视化

Fig. 8 Visualization of the comparison of the methods of this paper between the recognition results and the real results on the highway dataset and synthetic dataset

本文方法的优越性主要来自于两个方面，一方面是所提出的多通路体系结构能够有效地集成视觉特征、传输矩阵和场景深度信息，以进行能见度识别。另一方面，设计的多通路注意力融合模块可以自适应地选择对能见度识别重要的特征。而在下一节中，将通过消融实验来进一步评估这两个方面的有效性。

3.4 消融实验

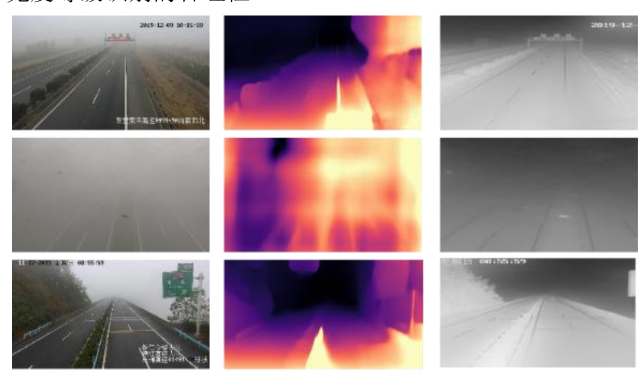
该网络具有多通路体系结构和自适应注意力融合的特点，在本节中，进一步进行了消融研究，以验证它们是否能够提高识别性能。首先为验证深度视觉通路中空洞卷积嵌入的有效性，构建了两种深度视觉通路结构：引入空洞卷积的残差网络(deep visual feature-1)与普通残差网络(deep visual feature-2)。并将两种结构的深度视觉特征通路与场景深度通路、传输矩阵通路、普通连接操作和多通路注意力融合作为设置选项，结合这些设置选项形成了五种简化网络模型(消融)。表 3 显示了五种消融方法和所提方法的识别精度。消融方法 1~4 使用的深度视觉通路均为嵌入空洞卷积的结构。消融方法 1 使用普通的连接操作来融合这三个流，消融方法 2、3 和 4 只使用其中某两个通路进行识别。通过比较所提出的 MSVP-Net 与消融方法 2-4，可以发现三种通路的组合具有最高的识别精度，通过比较所提出的 MSVP-Net 与消融方法 1，可以看出，自适应注意力融合也可以显著提高识别精度。消融方法 5 使用的深度视觉通路为普通残差网络，其余设置与所提出的方法 MSVP-Net 相同，通过比较消融方法 5 与所提方法 MSVP-Net 可以看出，由于雾的分布具有空间不均匀性，空洞卷积通过扩大感受野，提取多尺度特征能够更有利于学习雾的分布，助于能见度的识别。综上，这些消融结果充分证明了深度视觉通路结构、多通路组合与自适应融合的有效性。

表 3 本文方法与五种消融方法的识别精度

Tab. 3 The prediction accuracy of the proposed method and five ablation methods

Setting	MSVP-Net	ablation method 1	ablation method 2	ablation method 3	ablation method 4	ablation method 5
Deep Visual Feature-1	√	√	√	√	√	×
Deep Visual Feature-2	×	×	×	×	×	√
Scene Depth	√	√	√	×	×	√
Transmission Matrix	√	√	×	√	×	√
Concatenation Fusion	×	√	×	×	×	×
Attention Fusion	√	×	√	√	√	√
Accuracy	**84.68%**	81.70%	82.55%	78.30%	77.02%	83.83%

除此之外，本文给出了一些中间处理结果的可视化展示。图 9 展示了由场景深度通路和传输矩阵通路估计的高速公路监控图像的深度特征图和传输矩阵特征图。场景深度图中颜色较深的颜色，表示更大的场景深度。传输矩阵信息反映了在特定的空间位置上的传输速率,传输矩阵图中较亮的颜色，表示传播率较高。由于雾的影响，可以看到高速公路的场景深度逐渐缩小，同时传输速率也降低了，这在公路的远端尤为明显。这些结果也说明了利用场景深度和传输矩阵进行能见度等级识别的合理性。



(a)高速公路监控图像　　(b)估计的景深图　　(c)估计的传输矩阵图

图 9　三幅高速公路监控视频图像及其估计的景深图、传输矩阵图

Fig. 9　Three highway surveillance images and their estimated scene depth and transmission matrix

图 10 可视化展示了多通路注意力融合模块学习到的注意力图，第一行是原高速公路监控图像，第二行为学习到的注意力图并将其叠加于原始图像，其中黄色显示区域代表具有高权重的突出注意区域。从图 10 中看到，本文网络对于能见度识别的突出区域主要位于公路上，并沿着公路的延伸方向分布。对于能见度较低的场景，最突出的区域主要位于刚刚看不见消失在雾中的路段，对于能见度高的场景，在地平线上也会有突出的区域。从这些显著区域的分布来看，网络识别方式在某种程度上与人类观察能见度的方式相似。

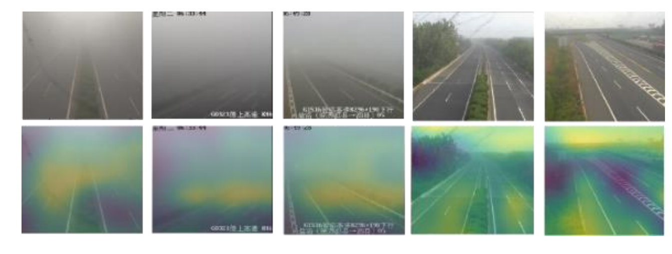


图 10　多通路注意力融合模块所学习的注意力图

Fig. 10　The visualization of the spatial attention maps learned by the multi-stream attention fusion module

4　结束语

本文提出了一种多通路深度融合网络，可以利用雾天情况下拍摄的监控图像识别高速公路能见度等级。该网络多通路体系结构的灵感来自于大气散射模型，具体而言，此网络

首先联合利用深度视觉特征通路、传输矩阵通路和场景深度通路进行特征学习。然后，设计了一个注意力融合模块，自适应地融合这三条通路得到的特征。最后，利用融合后的特征进行最终的能见度等级识别。实验结果表明，该网络在合成数据集与真实场景数据集上都具有良好的识别性能，充分验证了该方法的有效性和可行性。

参考文献：

[1] Hassaballah M, Kenk M A, Muhammad K, *et al.* Vehicle Detection and Tracking in Adverse Weather Using a Deep Learning Framework [J]. IEEE Trans on Intelligent Transportation Systems, 22 (7): 4230 – 4242, 2021.

[2] Hautière N, Aubert D, Dumont É, *et al.* Experimental validation of dedicated methods to in-vehicle estimation of atmospheric visibility distance [J]. IEEE Trans on Instrumentation and Measurement, 2008, 57 (10): 2218-2225.

[3] He K, Sun J, Tang X. Single image haze removal using dark channel prior [J]. IEEE Trans on pattern analysis and machine intelligence, 2010, 33 (12): 2341-2353.

[4] Li B, Peng X, Wang Z, *et al.* Aod-net: All-in-one dehazing network [C]// Proceedings of the IEEE international conference on computer vision. 2017: 4770-4778.

[5] Zhang J, Min X, Zhu Y, *et al.* HazDesNet: An End-to-End Network for Haze Density Prediction [J]. IEEE Trans on Intelligent Transportation Systems, 2020, online.

[6] 彭莉婷，李波．基于优化后透射率和半逆法的暗通道图像去雾方法 [J]．计算机应用研究, 2019, 36 (11): 3174-3178. (Peng Liting, Li Bo. Dark channel prior image dehazing method based on optimization transmission and semi-inverse algorithm [J]. Application Research of Computers, 2019, 36 (11): 3174-3178.)

[7] Nayar S K, Narasimhan S G. Vision in bad weather [C]// Proceedings of the 7th IEEE International Conference on Computer Vision. IEEE, 1999, 2: 820-827.

[8] Hautière N, Labayrade R, Aubert D. Real-time disparity contrast combination for onboard estimation of the visibility distance [J]. IEEE Trans on Intelligent Transportation Systems, 2006, 7 (2): 201-212.

[9] Hautiere N, Tarel J P, Lavenant J, *et al.* Automatic fog detection and estimation of visibility distance through use of an onboard camera [J]. Machine vision and applications, 2006, 17 (1): 8-20.

[10] Hautière N, Labayrade R, Aubert D. Estimation of the visibility distance by stereovision: A generic approach [J]. IEICE Transactions on information and systems, 2006, 89 (7): 2084-2091.

[11] Middleton W E K. Vision through the atmosphere [M]// Geophysik II/Geophysics II. Springer, Berlin, Heidelberg, 1957: 254-287.

[12] Gallen R, Cord A, Hautière N, *et al.* Nighttime visibility analysis and estimation method in the presence of dense fog [J]. IEEE Trans on Intelligent Transportation Systems, 2014, 16 (1): 310-320.

[13] Bronte S, Bergasa L M, Alcantarilla P F. Fog detection system based on computer vision techniques [C]// International IEEE Conference on Intelligent Transportation Systems. IEEE, 2009: 1-6.

[14] Mao J, Phommasak U, Watanabe S, *et al.* Detecting foggy images and estimating the haze degree factor [J]. Journal of Computer Science & Systems Biology, 2014, 7 (6): 226-228.

[15] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409. 1556, 2014.

[16] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[17] Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[18] 华夏，王新晴，马昭烨，等．复杂大交通场景弱小目标检测技术 [J]．计算机应用研究, 2019, 36 (11): 3486-3492. (Hua Xia, Wang Xinqing, Ma Zhaoye, *et al.* Detection of dim and small targets in complex large traffic scenes [J]. Application Research of Computers, 2019, 36 (11): 3486-3492.)

[19] Gunawan A A S, Prasetyo H, Werdiningsih I, *et al.* Inferring the level of visibility from hazy images [J]. International Journal of Business

Intelligence and Data Mining, 2020, 16 (2): 177-189.

[20] Liu F, Shen C, Lin G. Deep convolutional neural fields for depth estimation from a single image [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 5162-5170.

[21] Li Y, Huang J, Luo J. Using user generated online photos to estimate and monitor air pollution in major cities [C]// Proceedings of the 7th International Conference on Internet Multimedia Computing and Service. 2015: 1-5.

[22] Wofk D, Ma F, Yang T J, *et al.* Fastdepth: Fast monocular depth estimation on embedded systems [C]// International Conference on Robotics and Automation (ICRA) . IEEE, 2019: 6101-6108.

[23] Tarel J P, Hautiere N, Cord A, *et al.* Improved visibility of road scene images under heterogeneous fog [C]// IEEE Intelligent Vehicles Symposium. IEEE, 2010: 478-485.

[24] Tarel J P, Hautiere N, Caraffa L, *et al.* Vision enhancement in homogeneous and heterogeneous fog [J]. IEEE Intelligent Transportation Systems Magazine, 2012, 4 (2): 6-20.

[25] Silberman N, Hoiem D, Kohli P, *et al.* Indoor segmentation and support inference from rgbd images [C]// European conference on computer vision. Springer, Berlin, Heidelberg, 2012: 746-760.

[26] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv: 1412. 6980, 2014.

[27] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [J]. Advances in neural information processing systems, 2012, 25: 1097-1105.

[28] Dong H, Pan J, Xiang L, *et al.* Multi-scale boosted dehazing network with dense feature fusion [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2157-2167.